PB-217 570

DISCUSSION ON MOST PROBABLE NUMBERS IN SANITARY BACTERIOLOGY

V. G. MacKenzie

Sanitary Engineering Center
Cincinnati, Ohio

April 1949

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

DISCUSSION ON MOST PROBABLE NUMBERS IN SANITARY BACTERIOLOGY

LIBRARY
SANITARY ENGINEERING CENTER
CINCINNATI, OHIO
PB 217 570

(V. G. MacKenzie)

In sanitary bacteriology, the decimal dilution method has been widely used for estimating the densities of coliform bacteria in samples of water and other substances. It has appeared to be superior to the plate count for enumerating a specific organism in a mixed population by a test utilizing a selective medium of growth and particularly under conditions of low densities of the specific test organism. As generally used, the test has consisted of the inoculation of a number of tubes of a selective liquid medium with one or more decimal quantities or dilutions of the sample and observation of the inoculated tubes to determine the relative numbers in which the expected growth or reaction occurs. It is in effect, a determination of those tubes in which coliform bacteria were present in the portion of inoculated sample and those in which coliform bacteria were absent. From the results of relative numbers of positive and negative tubes, the number of coliform bacteria present in the sample must be estimated. Two methods for accomplishing this have been used: - The Phelps Index and the Most Probable Number Method.

The Phelps Index has had extensive use and is based on a very simple concept, namely, that the number of bacteria per unit of volume is the reciprocal of the amount of sample contained in the highest dilution showing growth. Thus, if growth is found in the tube containing 10 ml. of sample but not in that containing 1 ml, a density of 0.1 organisms per ml. is deduced. The objections to the Phelps method relate to (1) the inherent assumption of complete uniformity of the distribution of organisms in the universe sampled, (2) the neglect of the probability that positive portions of the highest dilution of the sample may contain more than one organism,

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

and (3) the difficulty of interpretation of results when growth is found in a dilution higher than one in which no growth is found, i.e. so called "skip-results."

It is easily seen that the neglect of the probability of a positive portion of the highest dilution of a sample containing more than one organism will result in a density estimate which is lower than the true value. For example, if 6 of 10 portions of 10 ml. each are found positive, the Phelps Index value would be 6 per 100 ml. If any of the six positive portions had contained more than one organism, an occurrence which is entirely consistent with the results and, indeed, likely according to the laws of chance, it is obvious that the total number of organisms in the 100 ml. tested would have exceeded six and the answer obtained by the Phelps Index Method would be too low.

The difficulty of dealing with "skip-results" by the Phelps Index Method derives basically from the conflict between the facts of experience and the Method's basic assumption of complete uniformity of distribution of organisms in the universe being sampled. Arbitrary procedures have had to be devised to cope with "skip-results" in the Phelps Index Method. Such procedures are fully outlined in Standard Methods.

The Most Probable Number Method of estimating coliform densities has come into almost universal use because it has offered a logical method of analysis which eliminated the basic objections to the Phelps procedure. The M.P.N. method was first proposed by McCrady in 1915 and since then has received the attention of many mathematicians. Because of the complexity of the mathematics associated with the computations involved, tables have been prepared for the use of bacteriologists using the method and these are in common use. This

has permitted an easy application of the M.P.N. method. This ease of use, however, has not contributed to a thorough understanding of the M.P.N. concept nor of the interpretation of the results reported.

The basic assumptions made in computing Most Probable Numbers are: First, that the bacteria are distributed at random in the material being sampled; Second, that the organisms exist independently of each other; and Third, that a single organism is sufficient to give a positive test. In relatively clean waters, these assumptions appear reasonable. In highly polluted waters, with many discrete particles, some question may be raised regarding them.

All of us have experience with the laws of chance or probability through the weather, playing cards, being stopped by traffic lights while driving, etc. In each such case on which we have adequate data the probability of a given result can be computed mathematically with more or less ease. For instance, with respect to a thoroughly shuffled deck of cards, the probability of not obtaining a heart on the first card exposed is 39/52 or 3/4. The probability of drawing one card from each of 2 thoroughly shuffled decks and not obtaining a heart in either instance would be 3/4 x 3/4; the probability of doing this with ten decks would be $(3/4)^{10}$. We have thus a relationship between the probability of drawing a heart from a universe as represented by a deck of cards and the concentration of hearts therein.

The relationship between the probability of a negative coliform fermentation tube test and an assumed density of organisms can be derived by a binomial series according to Poisson. The formula which applies is:

$$q = e^{-ax}$$

Where q = probability of a negative test

e = base of natural logarithms

a = volume of sample portion in ml (or other unit of volume)

x = average density of organisms per ml (or other unit of volume)

This formula is fairly easily derived, using the same reasoning as with the deck of cards.

If v = the unit volume of 1 organism

1/v = the number of such unit volumes in 1 ml

1/v -x = average number of such unit volumes in 1 ml. which are water.

If we take a sample of size 'v', the probability of getting a negative result is:

$$q = \frac{\text{no. of unit volumes of water}}{\text{total number of unit volumes}} = \frac{1/v - x}{1/v}$$

$$= 1 - vx$$

If a series of 'n' such samples of size 'v' are taken, the probability of all of them being negative is:

$$q = (1 - vx)^n$$

Or, if a 1 ml. portion is taken (n = 1/v)

$$q = (1 - vx)^{1/v}$$

This can be expanded by a McLaurin's series to

$$\text{In } q = -x - \frac{vx^2}{2} - \frac{v^2x^3}{3} - \frac{v^3x^4}{4} \quad \ldots\ldots$$

Since 'v' is a very small number indeed, the second and succeeding terms may be dropped and

$$q = e^{-x}$$

Or, if a sample portion of size 'a' ml is taken

$$q = e^{-ax}$$

It follows, of course, that the probability of a positive test is, of course, $1 - e^{-ax}$,

$$p = 1 - e^{-ax}$$

where P = probability of a positive test

The mathematics of M.P.N. analysis from this point on becomes more complicated. These simple equations, however, may be examined at this point since they have some application to the planting of tubes in a single dilution. If we have planted a series of tubes in a single dilution and observed the result, we have made an empirical measurement of the probability of the occurrence of a positive result. From the value of the probability thus obtained, we can compute the value of 'x' and this is the Most Probable Number. From an exercise of this kind, we can deduce several conclusions:

1. The estimated M.P.N. is dependent only on the percentage of positive (or negative) tubes. The figure is the same for 3 positives out of 5 as for 6 positives out of 10.
2. The precision of measurement is affected by the number of tubes used, since the number of possible answers varies directly with this quantity.
3. The answer with all tubes of the series positive, is indeterminately larger than the next smaller possible measurement.

The following table illustrates these points: -

M.P.N. for Various Numbers of 10-ml Tubes Positive

With 5 Tubes Planted		With 10 Tubes Planted	
No +	M.P.N.	No +	M.P.N.
		1	1.0
1	2.2	2	2.2
		3	3.5
2	5.1	4	5.1
		5	6.9
3	9.1	6	9.1
		7	12.0
4	16.0	8	16.0
		9	23.0
5	> 16.0	10	> 23.0

The precision of the M.P.N. determination is only lightly touched upon in the above table. This matter of precision is one of great importance and, unfortunately, is too often overlooked in practice. It is easy to look up the results of a test in a table and find the appropriate M.P.N. Too often there is a tendency to accept this as a figure of great accuracy and to forget that the measurement yardstick has been one with quite coarse divisions.

A number of investigators have attacked the problem of developing a measure of the precision which can be attached to an M.P.N. None of the methods developed really answer the question that the practical man would ask: "I have tested 5 - 10 ml portions of water for coliforms and have found one portion positive. This corresponds to an M.P.N. of 2.2 per 100 ml, but I would like to be sure what the error in this figure is likely to be."

There is no answer to this that does not assume something about the true density of organisms or make use of some prior knowledge to obtain such information. Most often it has been assumed that there was an equal probability of any density from 0 to infinity. Such an assumption obviously is not in accordance with the facts, particularly with respect to drinking waters.

One method of obtaining an approximate idea as to the precision involved in the M.P.N. procedure is to calculate the probability of the occurrence of a certain combination of positive and negative tubes. The plot of such probability as it varies with bacterial density gives one a visual conception of the precision. This is probably most easily done by the setting up of an equation for the total probability of occurrence of certain combinations of positive and negative tubes. Thus for a number of tubes, 'n', in a single dilution, the total probability is:

$$\left[e^{-ax} + (1 - e^{-ax}) \right]^n$$

You may recognize this as a combination of the probabilities for negative tubes and for positive tubes. As expanded, the first term would give the probability that no tubes would show growth, the second term that only one tube would show growth, etc. For any particular combination of positive and negative tubes, the probability of occurrence might be expressed by:

$$P = \frac{(r+m)!}{r! \cdot m!} (e^{-ax})^m (1 - e^{-ax})^r \quad \text{(Binomial law)}$$

Where r = number of tubes showing growth

m = number of tubes not showing growth

If we take this equation and compute P for various values of 'x', with 'r', 'm', and 'a' constant, we can obtain a plot of the type I have mentioned. For 5 - 10 ml portions planted in fermentation tubes with 1 tube positive the results are as follows:

Probability of Occurrence of 1 Positive Tube
Of 5 - 10 ml Portions for Various Values
of Coliform Density

Coliform Density	Probability
0.2/100 ml	0.091
0.5	0.200
1.0	0.319
1.5	0.383
2.0	0.407
2.2	0.410
3.0	0.390
4.0	0.333
5.0	0.266
6.0	0.205
7.0	0.153

You may notice that the maximum value of the probability as thus determined falls at a bacterial density of 2.2 per 100 ml. This, of course, is the M.P.N. However, you should also notice that considerable variation from this figure is, in this case, not materially less probable.

Another method of obtaining an approximate idea as to the reliability of M.P.N. figures is by assuming a coliform density and computing the relative frequency of various results to be expected. Let us assume that the average coliform density is 1.05 per 100 ml., the upper limit established for drinking water by the Public Health Service Standards and that we are again making tests with 5 - 10 ml tubes. The probability of a positive tube is $P = 1 - e^{-ax} = 1 - e^{-0.105} = 0.1$. That is, in the long run, 10 per cent of the 10 ml. tubes will be positive. The probability of getting various combinations of positive and negative tubes can be obtained from the previous equation outlined for this purpose. The following table shows the results for 5 tubes: (Table furnished showing this):-

Distribution of Results with 5 - 10 ml. Tubes
at Coliform Density of 1.05 per 100 ml.

No. of + Tubes	% of Samples	M.P.N.
0	59.05	< 2.2
1	32.80	2.2
2	7.29	5.1
3	0.81	9.1
4	0.05	16
5	0.00	> 16

Thus about 60 percent of the time the results would be entirely negative and less than 1 percent of the time would 3 or more tubes be positive. The Drinking Water Standards permit 5 percent of the samples to show 3 or more positive 10 ml. tubes. However, it should be noted that about 8 percent of the time we could expect results indicating an M.P.N. of 5.1 or more even though the true bacterial density was 1.05.

About 1930, the development of the method of confidence intervals made it possible to supply certain information about the precision of M.P.N. measurements. This method makes it possible to make certain overall statements regarding precision of results which, although not the exact statements in which we are interested, come very close to them. You have been furnished with a table showing the 50 to 95% confidence intervals for various results with 5 tubes of 10 ml. each. The 95 percent confidence intervals for those are as follows:-

No. of Positive Tubes	M.P.N. per 100 ml.	99% Confidence Interval per 100 ml.	
		Lower	Upper
0	0	0	9.2
1	2.2	.02	15.1
2	5.1	.36	22.4
3	9.1	1.12	34.1
4	16	2.5	84.8
5	∞	5.1	∞

One would like to be able to say that there is a 99% chance of being right if one states that given the result "one out of five positive" the density is between .02 and 15.1 per 100 ml. This is not possible. The proper statement can be made only about all six statements as a group. It may very well be that one of the statements will seldom be correct and the others will be correct more often than 99% of the time. All that we can say is that in a long series of tests of 5 tubes we will be right 99% of the time if we predict each time we get one of the above results that the coliform density lies in the interval corresponding to that result. Since what we usually are interested in is the confidence which can be placed in a given result rather than in a group of possible results the advantage of the method of confidence intervals is perhaps more apparent than real.

With respect to plantings in three decimal dilutions, you have also been furnished a table and a graph showing the approximate confidence limits of the M.P.N. determined from various numbers of tubes, (1,2,3,5 and 10 tubes) in three decimal dilutions.

The interpretation of these figures is as follows: When M.P.N. estimates are made on the basis of dilution tests using one tube in each of three decimal

dilutions you will be right 50% of the time if you say that the true bacterial density is between 33% and 186% of the M.P.N. If you have used 5 tubes in each dilution you could reduce this interval to from 64% to 139% of the M.P.N. and still be right 50% of the time. If a greater certainty were desired, say 95%, you would have to widen the interval to from 31% to 283%.

The actual determination of an M.P.N. has been reduced by now to finding the result in a table. Such tables are printed in Standard Methods for a few of the commoner combinations of tubes. Hoskin's paper, Reprint 1621 from Public Health Reports gives quite complete tables. You have been given copies of this reprint. Professor Thomas of the Harvard Engineering School has developed a very simple formula which is easy to remember and to use and gives very good approximations to the values computed by the rather laborious method necessary when dealing with the exact equations. This formula says:

$$\text{M.P.N.} = \frac{\text{No. of positive Tubes}}{\sqrt{(\text{Vol. in Positive Tubes})\ (\text{Total Vol. in all Tubes})}}$$

Reproduced from best available copy.

In this case the M.P.N. is in the same volume units as those on the right hand side of the equation. This equation is applicable to any number of tubes in any series of dilutions. The mimeographed sheet which you have been given shows two examples of the application of this formula. Although the values obtained from this formula should not be used except in the absence of the more accurate figures the formula does enable one to obtain approximate values very readily.

It is hoped that this discussion will bring to you an appreciation as to the accuracy to be attributed to M.P.N. values as they are ordinarily determined in sanitary bacteriology. It is important to be able to assign a definite number to the outcome of each test and the M.P.N. is the best estimate that has

been developed to date for this purpose. It is important, also, however, that a precision of measurement not be attached to such values which is not warranted by the facts.

Where bacterial densities are low as in the case of drinking water the dilution method will almost certainly continue to be the method used. However, in work with relatively polluted waters where coliform densities are of the order of 20 or more per ml. a more accurate estimate theoretically can be made by plate count than can be made from 10 or more tubes providing a suitable medium can be obtained which will be selective for coliforms. The standard error of a series of plate counts should be approximately equal to the square root of the mean count. Actual deviations encountered, of course, may depart from this norm. The plate count procedure, however, inherently has qualities of considerable mathematical accuracy. So far, apparently, no suitable medium for such plate count has been found. However, it is hoped that the search will continue because of the superior accuracy and ease of determinations which might be made possible thereby.

Reproduced from best available copy.